Journal of KIIT. Vol. 23, No. 11, pp. 171-180, Nov. 30, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.11.171

이미지-텍스트 통합 분석을 이용한 개인화 시각 콘텐츠 생성 시스템

박현지*[1], 김보민*[2], 노윤지*[3], 홍예빈*[4], 강지헌**

Personalized Visual Content Generation System using Image-Text Integrated Analysis

Hyunji Park*[1], Bomin Kim*[2], Yoonji Noh*[3], Yebin Hong*[4], and Jiheon Kang**

요 약

최근 MZ세대를 중심으로 이미지와 텍스트를 활용한 개인화된 일상 기록 문화가 확산됨에 따라, 외형과 감정을 기반으로 한 몰입형 시각 콘텐츠의 수요가 증가하고 있다. 본 연구에서는 사용자의 전신사진과 일기 분석을 통한 시각화 콘텐츠를 자동으로 생성하는 웹 기반 시스템을 제안한다. 해당 시스템은 이미지 스타일 인식과 텍스트 분석을 통해 외형 정보 및 키워드 추출한 후 이를 바탕으로 3D 캐릭터와 영상 일기 데이터를 생성한다. 이를 디지털 콘텐츠뿐 아니라 실물 3D 출력까지 확장하여 제안함으로써, 감정 기록을 보다 몰입감 있고 실체적인 경험으로 전환하는 가능성을 시사한다. 본 논문은 전체 시스템 아키텍처와 구성 요소를 통합적으로 제시하며, 객체 탐지 모델과 생성형 인공지능을 활용한 이미지-텍스트 통합 시스템의 활용 가능성을 제안한다.

Abstract

As image- and text-based personal journaling becomes popular among the Millennials and Gen Z, demand for appearance- and emotion-driven immersive content is growing. Our study propose a web-based system that automatically generates personalized visual content using users' full-body photos and diaries. The system extracts appearance and keywords through image-based style recognition and text analysis, then creates personalized 3D characters and video diaries. The output extends beyond digital visuals to real 3D prints, offering a more immersive and practical emotional recording experience. This paper presents the integrated system architecture and components, and proposes the image-text integration systems utilizing object detection models and generative artificial intelligence.

Keywords
personalized content generation, image detection, text analysis, sentiment analysis

* 덕성여자대학교 소프트웨어전공 학부과정
- ORCID[1]: https://orcid.org/0009-0004-3970-5777
- ORCID[2]: https://orcid.org/0009-0009-7511-2203
- ORCID[3]: https://orcid.org/0009-0008-5102-4519
- ORCID[4]: https://orcid.org/0009-0001-1334-8255
** 덕성여자대학교 소프트웨어전공 교수(교신저자)
- ORCID: https://orcid.org/0000-0001-5423-3895

· Received: Aug. 12, 2025, Revised: Oct. 21, 2025, Accepted: Oct. 24, 2025
· Corresponding Author: Jiheon Kang
Dept. of Software, Duksung Women's University, South Korea
Tel.: +82-2-901-8452, Email: jhkang@duksung.ac.kr

Ⅰ. 서 론

최근 MZ세대를 중심으로 일상을 자신만의 방식으로 기록하고, 이를 이미지와 텍스트 형태로 공유하는 문화가 확산되고 있다. ‘인생네컷’과 같은 셀프 포토 스튜디오는 단순히 사진 촬영을 넘어 개성을 표현하는 놀이 문화로 정착한 지 오래다. 또한 블로그 챌린지-이른바 ‘블챌’-의 유행은 기록을 통해 감정과 경험을 나누는 새로운 소통 수단의 트렌드를 보여준다. 이와 더불어 아바타 캐릭터 콘텐츠는 개인의 정체성과 감정 표현의 주체로 꾸준히 주목받고 있다. 과거 싸이월드의‘미니미’에서 시작해 최근에는‘제페토’,‘본디’ 등 메타버스 플랫폼의 부상과 함께 진화하고 있으며, 인공지능 기술이 대중화되면서 그 활용성이 더욱 확대되고 있다.

본 연구에서는 이러한 트렌드와 기술적 발전을 바탕으로 개인화 시각 콘텐츠 자동 생성 시스템의 설계 및 구현 방안을 제안한다. 이는 기존의 텍스트와 이미지가 분리된 일상 기록의 방식을 셀프 포토 스튜디오를 통해 통합하여 색다른 방식으로 시각화함으로써 사용자에게 몰입감 있는 경험을 제공한다.

Ⅱ. 관련 연구

2.1 셀프 포토 스튜디오의 확장성

‘인생네컷’과 같은 셀프 포토 스튜디오는 단순히 사진 촬영을 넘어, 사용자가 자신을 표현하고 기억을 시각적으로 기록하는 놀이 공간으로 자리 잡았다. 이는 사용자가 자기 정체성을 재구성하고 타인에게 전달하려는 의지를 뚜렷하게 나타낸다고 볼 수 있다[1]. 또한 다양한 콘텐츠 협업을 통한 촬영 프레임의 변화 등으로 셀프 포토 스튜디오의 확장 가능성을 엿볼 수 있다.

본 연구는 기존의 셀프 포토 스튜디오를 이미지 촬영과 텍스트 입력을 통합한 시스템으로 활용하며, 향후 관련 콘텐츠가 3D 프린트 분야로 확장될 가능성 또한 시사한다. 이는 정적인 이미지와 텍스트가 감정을 표현하는 실체화된 기록으로 연결되는 새로운 경험으로 작용할 수 있다.

2.2 YOLO 기반 의상 및 헤어스타일 분석

YOLO(You Only Look Once)는 한 번의 전처리 과정으로 객체 위치와 분류를 동시에 수행하는 One-stage 객체 탐지 기법으로서 실시간 외형 분석에서 유용하다. 기존의 Faster R-CNN(Faster Region-based Convolutional Neural Network)과 같은 Two-stage 모델보다도 가볍고 빠르며, 모바일 환경이나 클라이언트 기반 분석 환경에서 유리한 특성을 갖는다[2]. 최근 다양한 연구에서 YOLO를 패션 이미지의 의류 인식에 적용하여 그 성능과 실용성을 입증하였다.

YOLO 초기 버전부터 최근에 배포된 YOLOv11, YOLOv12까지 객체 탐지, 분류, 세그멘테이션 등 다양한 기능을 지원하며 YOLOv5 이후부터는 고성능 버전과 경량화된 세부 버전을 다양하게 선택적으로 사용할 수 있다. 본 연구에서 YOLO의 사용 목적은 객체 탐지이고, 탐지 대상이 되는 객체는 촬영된 사진 내 의상, 헤어 부분이여서 비교적 작은 객체에 해당되지 않아, 성능 개선과 경량과 관점에서 효율성을 비교할 수 있는 YOLOv5를 선택하여 본 연구에 적용 가능한 세부 모델을 비교 검증하고 활용하였다.

특히 YOLOv5는 경량화된 구조 설계, Mosaic 기반 데이터 증강, 다중 스케일 피처를 통합하는 FPN(Feature Pyramid Network) 및 PAN(Path Aggregation Network) 구조를 기반으로 다양한 크기와 형태의 객체를 높은 정확도로 탐지한다[3].

본 연구에서는 YOLOv5를 사용하여 사용자의 전신사진 속 헤어스타일과 의상을 실시간으로 분석하고, 이를 3D 캐릭터의 외형에 반영하는 데 핵심 역할을 할 수 있도록 적용하였다. 또한 YOLOv5의 다양한 버전의 경량화 관점, 성능 관점의 모델 비교를 통해 실시간 탐지의 반응 속도와 전체 시스템의 성능을 동시에 향상시킬 수 있도록 하였다.

YOLO의 경우 다양한 버전이 존재하며, 학습데이터의 종류, 탐지하고자 하는 객체의 형태에 따라 탐지 성능 관점 배포 시 모델 효율성 관점에서 적합한 YOLO버전 및 세부 모델을 적용할 수 있다.

2.3 LLM 기반 콘텐츠 생성 시스템

최근 LLM(Large Language Model)을 활용한 콘텐츠 생성 시스템 연구가 활발히 이루어지고 있다. 선행 연구에서는 LangChain과 RAG(Retrieval-Augmented Generation) 기반 구조를 적용하여 논문 요약, 리뷰 작성, 발표 자료 자동 생성을 지원하는 연구 지원 시스템이 제안되었다[4]. 이는 LLM을 통해 텍스트를 요약하고, 핵심 키워드를 추출하여 후속 자료 생성까지 연결하는 콘텐츠 생성의 한 사례라 할 수 있다. 텍스트 기반으로 개인화된 음성, 영상 콘텐츠를 생성해주는 시스템과[5] 교육현장에서부터 숏폼 제작까지 텍스트, 이미지, 오디오, 비디오 기반의 다양한 멀티모달 입력을 활용하여 생성형 AI 기반 다양한 콘텐츠 제작 사례들이 증가하고 있다[6][7].

본 연구 역시 이와 유사하게 영상분석과 LLM을 활용하여 텍스트 입력을 기반으로 한 콘텐츠 생성을 수행한다. 사용자의 일기 텍스트를 요약하고 감정·장소·사물 키워드를 추출하여 이미지 분석 데이터와 결합함으로써 개인화된 영상 일기를 생성한다. 즉, LLM 기반 텍스트 처리 과정을 한가지 핵심 축으로 하여 새로운 형태의 개인화 시각 콘텐츠 생성을 실현한다.

III. 시스템 설계 및 구현

본 시스템은 사용자의 전신사진과 일기를 분석하여 사용자의 의상 및 헤어 착장 특성, 장소, 감정, 사물 정보를 포함한 영상 일기를 자동 생성하고, 생성된 영상 일기 내 3D 캐릭터를 실사화하는 것을 목표로 한다. 이미지 분석을 통해 헤어스타일과 의상의 종류 및 색상 정보를 추출하고 이를 Unity로 연동될 3D 캐릭터 형상에 반영한다. 또한 텍스트 분석을 통해 감정, 장소, 사물 키워드를 추출하여 감정은 3D 캐릭터의 표정과 행동으로 장소는 배경 이미지로 사물 키워드는 이모지로 각각 맵핑한다. 각각 3D 캐릭터의 표정 및 모션 적용, 배경 이미지 생성, 영상 일기 속 이모지 적용을 통해 최종 결과물을 사용자에게 제공한다.

전체 시스템은 웹 기반으로 구현되었으며, 시스템의 구성은 그림 1과 같고 사진 촬영, 텍스트 입력을 시작으로 영상분석 결과와 텍스트 분석 결과를 Unity 파이프라인을 거쳐 최종 결과물을 출력하는 동작 흐름은 그림 2과 같다.

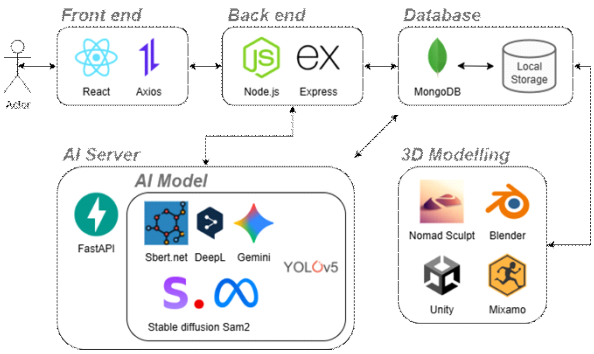

그림 1. 시스템 구성도
Fig. 1. System architecture

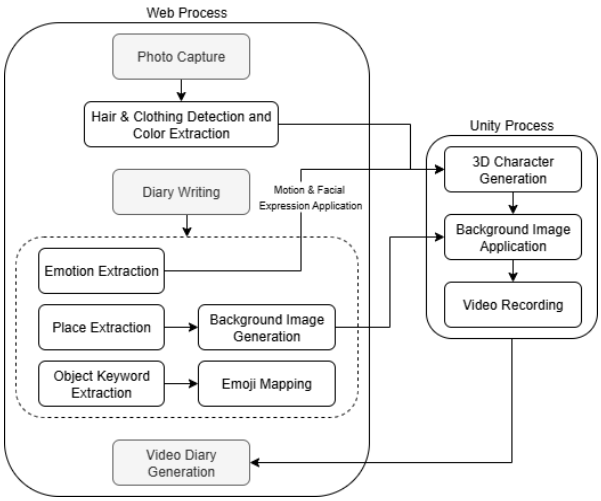

그림 2. 시스템 동작 흐름
Fig. 2. System operation flow

본 연구에서는 사용자로부터 획득한 감정 및 객체 분석 결과를 기반으로 Unity 엔진을 활용하여 캐릭터의 외형, 표정, 배경 및 모션을 동적으로 구성하고, 최종적으로 영상 일기를 생성하는 자동화 시스템을 구현하였다. 해당 시스템은 Unity의 MonoBehaviour 기반 구조를 바탕으로 동작하며, 외부 서버와의 비동기 통신을 통해 실시간 데이터를 수신하고 이를 바탕으로 3차원 환경 내 시각 요소를 자동으로 설정하는 절차를 포함한다.

시스템은 Unity의 Coroutine 기능과 Unity WebRequest API를 활용하여 주기적으로 감정 분석 식별자를 요청하고, 새로운 ID가 수신되면 관련된

감정 및 객체 분석 결과를 FastAPI 기반 외부 서버로부터 획득한다. 수신된 결과는 JSON 형식으로 파싱되어 Unity 내부의 각 시각 요소 구성에 반영된다.

3.1 YOLO 객체인식 모델 기반 스타일 분류

YOLO 모델은 학습을 위해 객체가 포함된 이미지와 각 객체의 위치를 정의하는 bounding box 좌표, 해당 객체에 대한 클래스 정보를 포함하는 라벨 파일을 요구한다. 이를 위해 헤어스타일과 패션스타일에 대한 클래스 정의를 명확히 하고 데이터세트 수집 및 전처리를 통해 객체 인식 모델 학습에 필요한 구조를 구축하였다.

3.1.1 헤어스타일 분류

1) 데이터 수집 및 전처리

AI Hub에서 제공하는 약 50만 장의 한국인 헤어스타일 이미지 중 총 67,460장을 선별하여 학습용 데이터세트를 구성하였다[8].

이 중 54,986장은 학습용(train), 12,474장은 검증용(val)으로 분할 하였다. 각 이미지는 정의된 클래스 기준인 앞머리 유무와 머리 길이에 따라 라벨링되었으며, 이후 클래스 간 불균형 문제를 최소화하기 위해 전체 데이터에서 각 클래스별 이미지 수를 유사한 수준으로 조정하였다. 조정된 학습 및 검증 데이터세트 분포는 표 1과 같다.

표 1. 헤어스타일 분류를 위한 데이터세트 구성
Table 1. Dataset for hairstyle classification

Class	Train	Val
no_bang_long	8,984	1,972
no_bang_mid	8,813	2,013
no_bang_short	9,721	2,244
bang_long	9,780	2,177
bang_mid	7,885	2,012
bang_short	9,803	2,056
total	54,986	12,474

2) 모델 학습

헤어스타일 탐지 모델 학습은 YOLOv5s 구조를 기반으로 수행하였다. 학습의 안정성과 수렴 성능을 고려하여 다양한 하이퍼파라미터 조정 실험을 병행하였다.

초기 학습에서는 모델의 학습 안정성 확보를 위해 손실 함수의 가중치를 조정하였으며, 객체 탐지의 정확도 향상을 위해 클래스 예측(cls)과 객체 존재 유무(obj) 항목의 가중치를 높게 설정하였다. 또한, 최적화 알고리즘은 일반적인 SGD(Stochastic Gradient Descent)를 채택하고, 학습률 스케줄링에는 Cosine annealing 방식을 적용하여 학습 후반부의 과도한 발산을 방지하였다. 주요 하이퍼파라미터는 표 2에 정리되어 있다.

학습 성능은 반복적인 튜닝을 통해 지속적으로 개선되었으며, mAP@0.5(mean Average Precision @IoU=0.5) 기준으로 초기 0.553에서 0.685로 향상되었고, 최종적으로는 0.733을 달성하였다. 이때 정밀도(Precision)는 0.86, 재현율(Recall)은 0.84를 기록하여 두 지표 간 균형도 양호한 것으로 나타났다.

표 2. 헤어스타일 분류를 위한 학습 파라미터 설정
Table 2. Hyper parameter settings for hairstyle detection model training

Image size	Batch	Epoch	lr0	lrf
640X640	16	100	0.01	0.1

3.1.2 패션 스타일 분류

1) 데이터 수집 및 전처리

AI Hub에서 제공하는 K-Fashion 이미지 데이터세트를 활용하였다[9]. 해당 데이터세트는 실제 착장 이미지 120만 건으로 구성되어 있으며 각 이미지에는 bounding box 좌표, 세부 의류 속성, 스타일 정보 등이 제공된다.

데이터세트는 의류의 category 필드에 포함된 세부 속성에 따라 상의 13종, 하의 7종, 아우터 4종의 총 24개의 클래스로 분류하여 약 34만 건의 적합한 이미지를 필터링하여 8:2 비율로 분할하여 학습 및 검증 데이터로 활용하였다.

2) 모델 학습

YOLOv5 세부모델의 구조적 차이와 데이터 불

균형 처리 유무에 따른 탐지 성능의 영향을 분석하기 위해 총 4가지 실험을 설계하였다. 비교 대상은 모델의 크기(YOLOv5s vs YOLOv5m)와 하이퍼파라미터 조정 여부이다. 모든 실험은 해상도 512×512, 배치 크기 12, 초기 가중치 YOLOv5s 조건하에 수행되었으며, Nvidia RTX3080Ti GPU 1개 환경에서 진행되었다.

먼저, Base-s와 Base-m 실험은 각각 YOLOv5s와 YOLOv5m 모델을 사용하였으며, 기본 파라미터 설정 하에서 데이터 불균형에 대한 별도 조정 없이 학습을 진행하였다. 이를 통해 모델 크기 자체가 탐지 성능에 미치는 영향을 비교하고자 하였다.

반면, Balance-s와 Balance-m 실험은 동일한 모델 구조를 사용하되, 데이터 불균형을 보완하기 위해 하이퍼파라미터를 조정하였다. 구체적으로는 클래스 불균형으로 인해 분류 오류가 자주 발생하는 소수 클래스를 더 잘 학습하도록 cls 파라미터를 0.3에서 1.5로 증가시켰고, 데이터 다양성을 확보하기 위해 copy paste 및 degrees 파라미터를 각각 0.2, 5.0으로 조정하였다. 적용모델별 검증 데이터에 대한 결과는 표 3과 같다.

표 3. 패션 스타일 분류 결과
Table 3. Train result of fashion style classification

	Precision	Recall	mAP@0.5
Base-s	0.711	0.795	0.787
Base-m	0.749	0.844	0.83
Balance-s	0.786	0.824	0.86
Balance-m	0.828	0.874	0.903

실험 결과, YOLOv5m 모델이 YOLOv5s에 비해 전반적으로 더 우수한 성능을 보였으나, 연산량이 더 크다는 단점이 존재하였다. 특히 mAP 측면에서 뚜렷한 차이를 보였으며, 이는 더 깊고 복잡한 구조가 더 많은 특징을 포착할 수 있기 때문으로 해석된다. 또한 데이터 불균형을 완화한 Balance 실험군에서는 Base 실험군 대비 성능 개선이 확인되었으며, 이는 특히 소수 클래스에서의 탐지 성능(mAP) 향상에 긍정적으로 작용하였다.

그림 3과 같이 confusion matrix 시각화 결과, 대부분의 클래스에서 안정적인 탐지 정확도를 보였으며 소수 클래스에서도 false positive와 false negative가 크게 줄어든 것이 확인되었다. 이는 데이터 불균형을 완화하기 위한 파라미터 조정의 효과로 해석할 수 있다.

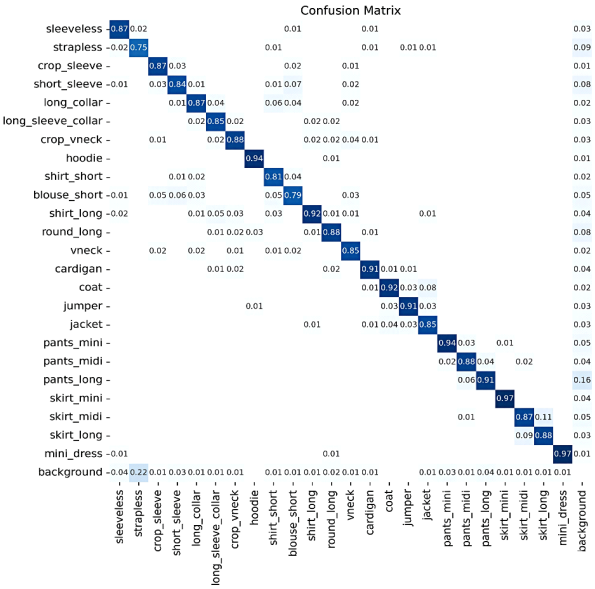

그림 3. Balance-m 모델 클래스별 학습 결과
Fig. 3. Confusion matrix of balance-m model

3.2 언어모델 기반 일기 분석

3.2.1 감정 분석 및 분류

감정 분석의 전체 흐름은 두 단계로 나뉜다. 첫 번째 단계에서는 Google의 'gemma-3-12b-it'를 활용하여 일기 텍스트로부터 감정 정보를 추출한다. 출력 결과는 주요 감정 한 개와 세부 감정을 최대 세 개까지 감정 강도와 함께 배열 형태로 제공하도록 구성된다.

두 번째 단계에서는 추출된 감정 데이터를 바탕으로 SBERT(Sentence-BERT, Bidirectional Encoder Representations from Transformers) 기반 유사도 분석을 활용하여 정형화된 감정을 도출한다. 그림 4와 같이 Russell의 Circumplex Model을 기반으로 2차원 감정 공간을 네 영역으로 분할하고, 각 영역별 대표 감정 단어를 정의하였다[10][11]. 추출된 감정 데이터를 SBERT 모델을 통해 임베딩하고, 각 대표 감정 영역의 중심 벡터와 코사인 유사도를 계산한다. 이때 감정 강도를 가중치로 반영하여, 강도가 높은 감정 표현이 분석 결과에 더 크게 기여하도록 하였다.

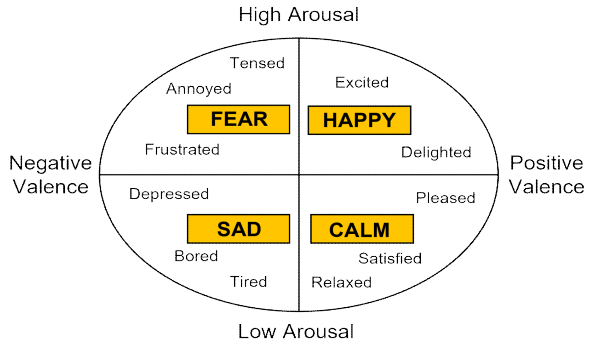

그림 4. Russell 감정 분류 모델
Fig. 4. Russell's circumplex model

최종적으로 선정된 감정분석 결과는 그림 5과 같이 Unity와 연계된 캐릭터의 표정 및 모션을 결정하며, 실제 시스템에 적용 시 영상 일기 내 캐릭터의 움직임 표시 시간을 고려하여, 기본값은 최종감정 1개에 대한 표정 및 모션 상태가 캐릭터에 반복 표현될 수 있도록 개발 시스템 GUI에 적용하였고, 필요시 2개 이상의 감정이 반복적으로 일정 시간 동안 표시될 수 있도록 구현하였다.

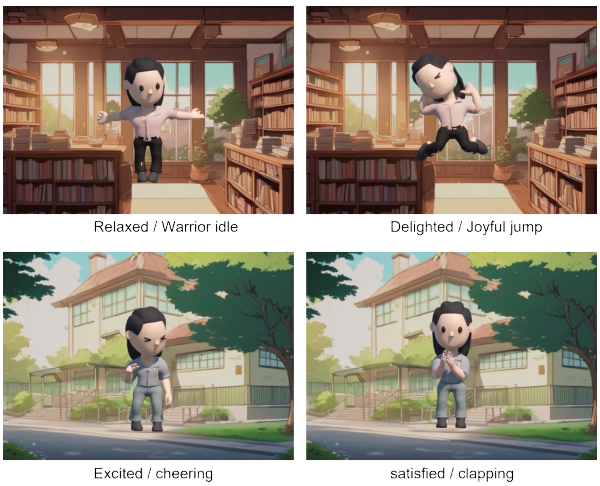

그림 5. 감정분류에 따른 캐릭터 표정 및 모션
Fig. 5. Character expressions and motion based on emotion analysis

3.2.2 코사인 유사도 기반 키워드-이모지 매핑

분석된 최종감정 1개와 일기 텍스트 내 사물 키워드 5개를 추출하여 그림 6와 같이 총 6개의 단어에 대응하는 적절한 이모지를 자동으로 선택한다. 사물 키워드 추출은 역시 Google의 'gemma-3-12b-it'를 활용하며, 사전 학습된 문장 임베딩 모델인 Sentence Transformer의 'all-MiniLM-L6-v2'를 통해 입력된 자연어를 고차원 임베딩 벡터로 변환하고, 사전 정의된 이모지 리스트와 유사도를 계산하여 상위 1개의 결과만을 최종 이모지로 결정한다.

문장 임베딩 모델이 학습한 언어와 호환성을 위해 Deep Translator 라이브러리의 Google Translator 번역 모델을 도입하였으며, 유사도 계산에는 PyTorch 기반의 코사인 유사도 함수인 pytorch_cos_sim을 사용하였다.

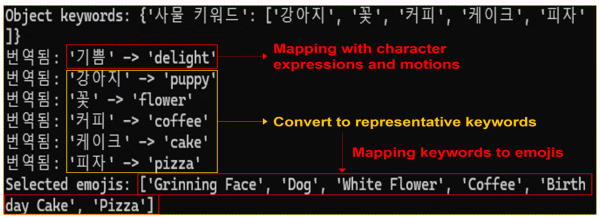

그림 6. 키워드-감정/이모지 매핑
Fig. 6. Keyword-emoji mapping

3.3 배경 이미지 생성

배경 이미지 생성은 Stable Diffusion v1.6 모델을 활용하였다[12]. 그림 7와 같이 일기 분석에서 추출한 장소 키워드를 기반으로 배경 이미지를 실시간으로 생성한다. 프롬프트는 2D 디지털 애니메이션 일러스트 콘셉트를 기반으로, 부드러운 셀 셰이딩과 수채화 느낌을 조화롭게 포함하는 스타일로 구성하였다. 생성 파라미터로는 CFG(Class-Free Guidance) 스케일 7, FAST_BLUE 클립 가이드 프리셋, 512×768 해상도, 30 스텝의 샘플링을 사용하여, 최적의 품질과 처리 속도의 균형을 달성하였다.

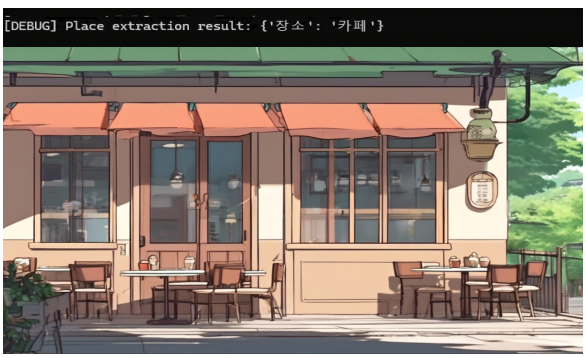

그림 7. 장소 키워드 기반 배경 이미지 생성
Fig. 7. Background image creation

3.4 영상 일기 제작

사용자의 전신사진을 앞서 설명한 흐름에 따라 분석한 결과를 기반으로 그림 8와 같이 맞춤형 캐릭터를 생성한다.

그림 8. 사용자 맞춤형 3차원 캐릭터 생성
Fig. 8. Personalized 3D character creation

3D 캐릭터의 기본 바디는 Adobe Mixamo에서 자동 리깅을 통해 FBX(Film box) 파일을 기반으로 구성되었으며, 각 헤어 및 의류 에셋은 해당 캐릭터의 본 구조에 적합하도록 리깅 및 자동 웨이트 부여 방식과 수동 웨이트 조정 방식을 병행하여 적용하였다. 리깅이 완료된 에셋은 Unity C# 스크립트를 활용해 의상 적용 작업을 자동으로 수행하여 최종 사용자 맞춤형 캐릭터를 생성하도록 구현하였다.

캐릭터 모션은 Unity의 Animator 시스템을 활용하여 애니메이션 시퀀스를 재생하여 구현된다. 감정-표정-모션 매핑 정보를 적용하여 모션 클립이 실행됨과 동시에 눈, 눈썹 등 특정 얼굴 구성요소가 개별적으로 활성화되어 복합적인 감정 표현을 구현한다.

배경의 경우, 텍스처를 평면 오브젝트의 Material에 동적으로 할당하여 시각적 배경을 구성한다. 캐릭터는 배경 평면 기준으로 Z축 방향 -3의 적절한 거리로 자동 조정되어 화면 내 최적의 배치를 확보한다.

최종 영상 녹화는 Unity Editor에서 제공하는 Recorder 패키지를 활용하여 구현되었다. MP4 형식의 비디오 파일로 저장되며, 이에 더해 캐릭터의 3D 메쉬 구조는 STL(Stereolithography) 파일로 추출되어 사용자에게 3D 프린터로 물리적 형상을 소장할 가능성을 제시한다.

생성된 영상과 일기 텍스트, 이모지 등을 다시 웹 페이지에서 조합하여 그림 9과 같은 완성된 영상 일기를 사용자에게 제공한다. 추가적으로 LLM을 활용한 일기 텍스트 요약 및 감정 분석 리포트를 함께 보여준다. 또한 해당 페이지의 캡처 이미지 및 3D 캐릭터의 STL 파일을 사용자에게 이메일로 전송할 수 있도록 하였다. 그림 10은 3D 캐릭터의 STL파일을 이용하여 3D 프린터로 출력한 결과물을 나타낸다.

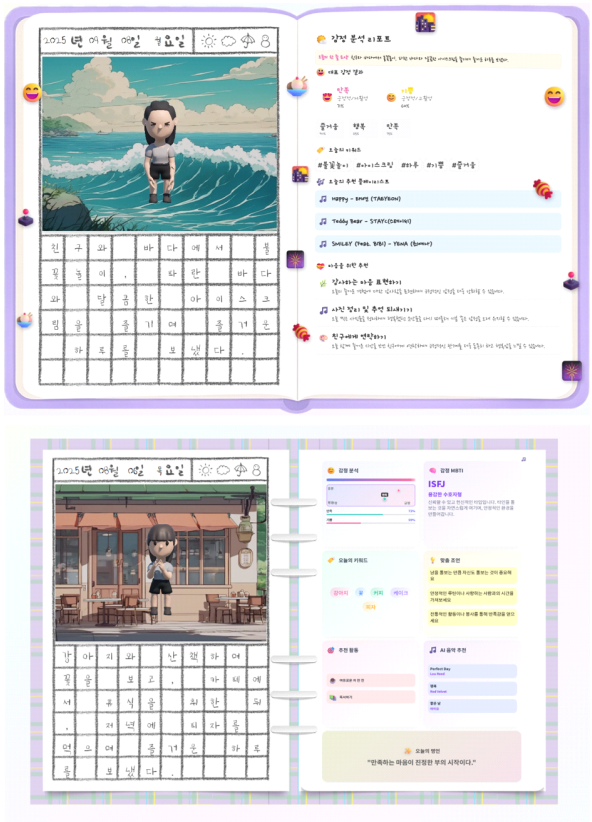

그림 9. 개인화 영상 일기 최종 결과물
Fig. 9. Personalized video diary results

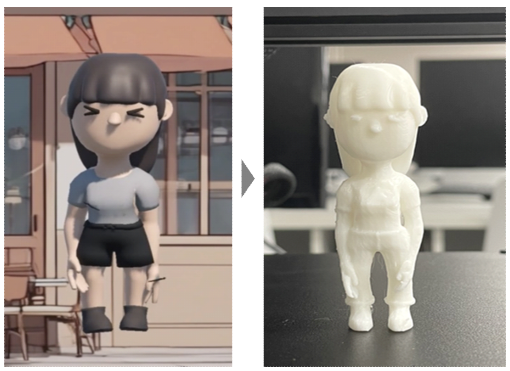

그림 10. 모델링 객체의 3D 프린터 출력 결과물
Fig. 10. 3D modeling object using a 3D printer

3.5 시스템 통합 성능

시스템의 통합 성능을 측정한 결과, 각 단계별 소요 시간은 다음과 같다. 이미지 분석 과정은 약 5~8초 이내에서 완료되었으며, 텍스트 분석 및 배경 이미지 생성 과정은 20초 이내의 시간이 소요되었다. 또한, Unity 기반 캐릭터 동작 및 영상 녹화 과정은 약 10초 내외로 진행되었다. 따라서 사용자의 입력 시간을 제외하면, 하나의 콘텐츠가 생성되어 결과를 확인하기까지 총 40초 이내에 완료되는 것을 확인할 수 있었다.

Ⅳ. 결론 및 향후 과제

본 시스템은 사용자의 전신사진 촬영과 일기 텍스트 입력을 바탕으로 개인화된 캐릭터를 생성하고, 이를 영상 일기 형태로 기록할 수 있도록 구성된 웹 기반의 자동 콘텐츠 생성 시스템이다. 즉, YOLO 모델과 LLM을 통한 데이터 분석을 토대로 Unity의 3D 캐릭터에 다양한 헤어스타일과 의상, 표정과 모션을 적용하며 Stable Diffusion을 활용한 배경 이미지의 결합과 함께 하나의 개인화된 영상 일기 콘텐츠를 제공한다. 이를 셀프 포토 스튜디오라는 공간에서 통합하여 일상을 더욱 특별하게 기록하는 경험을 제공하는 것이 본 연구의 핵심 목표이다.

다만 본 연구에는 몇 가지 한계점이 존재한다. 의상과 헤어스타일의 종류가 정확히 식별되고, 이를 캐릭터 생성 환경에 반영하는 과정이 본 연구의 핵심이다. 그러나 본 연구에서 활용한 AI Hub의 데이터세트는 실제로 더 많은 의상과 헤어스타일 클래스를 포함하고 있음에도 이를 모두 활용하지 못하였는데, 각 항목을 시각적으로 구현하는 데 어려움이 있어 Unity Asset이 제공하는 제한된 범위에서만 활용할 수 있었기 때문이다. 따라서 현 단계에서는 식별 및 적용 가능한 클래스의 종류가 제한적이라는 한계가 있으며, 추후 시각적 구현의 한계가 개선된다면 더욱 다양한 의상과 헤어스타일을 반영할 수 있을 것이다. 또한, 실제 사용자들에 대한 평가를 기반으로 UI 및 배경, 캐릭터에 대한 다양하고 사실적 표현 방법에 대한 개선이 필요할 것으로 판단된다.

향후 과제로 셀프 포토 스튜디오를 기반으로 한 시스템의 특성상 디지털 콘텐츠에 머무르지 않고 3D 프린터 출력으로 확장하는 방안을 제시한다. 현재 시스템은 3D 캐릭터 모델을 STL 파일로 자동 추출하여 사용자에게 메일을 전송하는 방식으로 실물화의 가능성을 제공하고 있다. 그러나 3D 프린터의 출력 속도가 개선되어 단시간에 결과물을 제작할 수 있는 수준에 도달한다면 이를 자동화하는 것도 가능하다. 이는 단순한 디지털 시각화에 그치지 않고, 사용자의 하루를 입체적으로 기억하고 보존할 수 있는 새로운 형태의 감정 기록 방식을 제시하는 시도로써 의미를 가질 것이다.

References

[1] S. Y. Hong and S. I. Kim, "Study on the content scalability of Generation Z self photo studio", Journal of Next-generation Convergence Information Services Technology, Vol. 12, No. 4, pp. 503-513, Aug. 2023. https://doi.org/10.29056/jncist.2023.08.08.

[2] J. Y. Kim, "A comparative study on the characteristics of each version of object detection model YOLO", Proc. of the Korean Society of Computer Information Conference, Daejeon, South Korea, pp. 75-78, Jul. 2023.

[3] Y. H. Chang and Y. Y. Zhang, "Deep learning for Clothing Style Recognition using YOLOv5", Micromachines, Vol. 13, No. 10, pp. 1678, Oct. 2022. https://doi.org/10.3390/mi13101678.

[4] C. Jeong, "Generative AI service implementation using LLM application architecture: based on RAG model and LangChain framework", Journal of Intelligence and Information Systems, Vol. 19, No. 4, pp. 129-164, Dec. 2023. https://doi.org/10.13088/jiis.2023.29.4.129.

[5] J. M. Lee and S. J. Kim "Development of a Text-based Personal Content Creation System using Deep Learning", The Journal of Korean Institute of Information Technology, Vol. 21, No. 12, pp.

67-75, Dec. 2023. https://doi.org/10.14801/jkiit.2023.21.12.67.

[6] S. Kim, "Case Study on Content Creation Program using Multimodal-based Generative AI in University Education", The Journal of the Korea Contents Association, Vol. 24, No. 11, pp. 609-617. Nov. 2024 https://doi.org/10.5392/JKCA.2024.24.11.609.

[7] D. H. Jo, M. S. Kim, S. M. Kim, and G. D. Kim, "A Generative AI-based Next-generation Short-form Creation Platform. Journal of Digital Contents Society, Vol. 25, No. 7, pp. 1701-1713, Jul. 2024 https://doi.org/10.9728/dcs.2024.25.7.1701.

[8] AI-Hub, Korean Hairstyle Image Dataset, https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=85. [accessed: Dec. 12, 2024]

[9] AI-Hub, K-Fashion Image Dataset, https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=51. [accessed: Apr. 16, 2025]

[10] J. A. Russell, "A circumplex model of affect", Journal of Personality and Social Psychology, Vol. 39, No. 6, pp. 1161-1178, 1980. https://doi.org/10.1037/h0077714.

[11] N. M. Nor and S. H. S. Salleh, "Correlation between precursor emotion and human stress by using EEG signals", ARPN Journal of Engineering and Applied Sciences, Vol. 10, No. 23, pp. 17881-17889, Dec. 2015.

[12] D. H. Kwon, "Analysis of Prompt Elements and Use Cases in Image-Generating AI: Focusing on Midjourney, Stable Diffusion, Firefly, DALL·E", Journal of Digital Contents Society, Vol. 25, No. 2, pp. 341-354, Jan. 2024. https://doi.org/10.9728/dcs.2024.25.2.341.

저자소개

박 현 지 (Hyunji Park)

2022년 3월 ~ 현재 : 덕성여자대학교 소프트웨어전공 학사과정

관심분야 : 웹 프로그래밍, 서버 프로그래밍, 인공지능, 사물인터넷

김 보 민 (Bomin Kim)

2022년 3월 ~ 현재 : 덕성여자대학교 소프트웨어전공 학사과정

관심분야 : 서버 프로그래밍, 모바일 프로그래밍, 인공지능, 빅데이터

노 윤 지 (Yoonji Noh)

2022년 3월 ~ 현재 : 덕성여자대학교 소프트웨어전공 학사과정

관심분야 : 웹 프로그래밍, 서버 프로그래밍, 인공지능, 빅데이터

홍 예 빈 (Yebin Hong)

2022년 3월 ~ 현재 : 덕성여자대학교 소프트웨어전공 학사과정

관심분야 : 서버 프로그래밍, 빅데이터, 인공지능

강 지 헌 (Jiheon Kang)

2007년 2월 : 강원대학교
전자통신공학(공학사)
2013년 8월 : 고려대학교
전기전자공학(공학석사)
2019년 2월 : 고려대학교
전기전자공학(공학박사)
20219년 3월 ~ 2021년 2월 :
SK텔레콤 ICT기술센터 매니저
2021년 3월 ~ 현재 : 덕성여자대학교 소프트웨어전공
조교수